Data Entry

with Sheet and SQL

[Your Full Name Here]

DATA ANALYST

Introduction

The dataset contains 1500 rows with various issues such as:

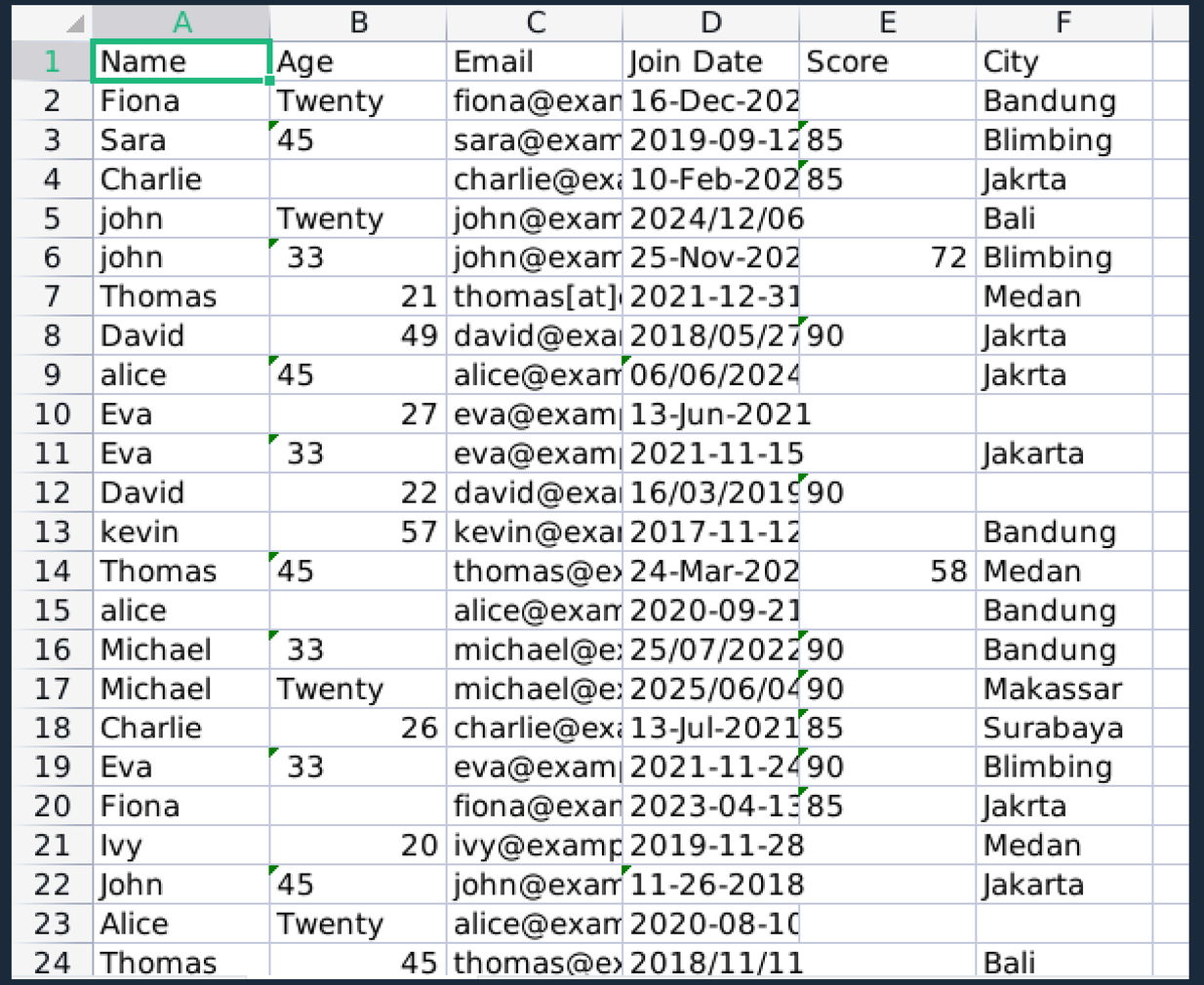
- The dataset contains 1500 rows with various issues such as:
- Missing values (age, score, city)
- Mixed date formats
- Duplicate names
- Inconsistent text formats
- This presentation shows step-by-step how the data was cleaned.

Data cleaning and verification are completed within one day, as the use of SQL and Excel enables a fast and efficient workflow.

Disclaimer: Name of Project is disguised

Beginning

First, the data had many issues, such as emails with no dormant scores, empty fields, messy spacing, and inconsistent use of upper and lower case letters, as shown in this image.



	A	B	C	D	E	F
1	Name	Age	Email	Join Date	Score	City
2	Fiona	Twenty	fiona@exar	16-Dec-202		Bandung
3	Sara	45	sara@exam	2019-09-12	85	Blimbing
4	Charlie		charlie@exa	10-Feb-202	85	Jakrta
5	john	Twenty	john@exam	2024/12/06		Bali
6	john	33	john@exam	25-Nov-202	72	Blimbing
7	Thomas	21	thomas[at]e	2021-12-31		Medan
8	David	49	david@exa	2018/05/27	90	Jakrta
9	alice	45	alice@exam	06/06/2024		Jakrta
10	Eva	27	eva@exam	13-Jun-2021		
11	Eva	33	eva@exam	2021-11-15		Jakarta
12	David	22	david@exa	16/03/2019	90	
13	kevin	57	kevin@exa	2017-11-12		Bandung
14	Thomas	45	thomas@ex	24-Mar-202	58	Medan
15	alice		alice@exam	2020-09-21		Bandung
16	Michael	33	michael@e	25/07/2022	90	Bandung
17	Michael	Twenty	michael@e	2025/06/04	90	Makassar
18	Charlie	26	charlie@exa	13-Jul-2021	85	Surabaya
19	Eva	33	eva@exam	2021-11-24	90	Blimbing
20	Fiona		fiona@exar	2023-04-13	85	Jakrta
21	Ivy	20	ivy@examp	2019-11-28		Medan
22	John	45	john@exam	11-26-2018		Jakarta
23	Alice	Twenty	alice@exam	2020-08-10		
24	Thomas	45	thomas@ex	2018/11/11		Bali

Find and Replace

Second, use Find and Replace to convert month names to numbers:

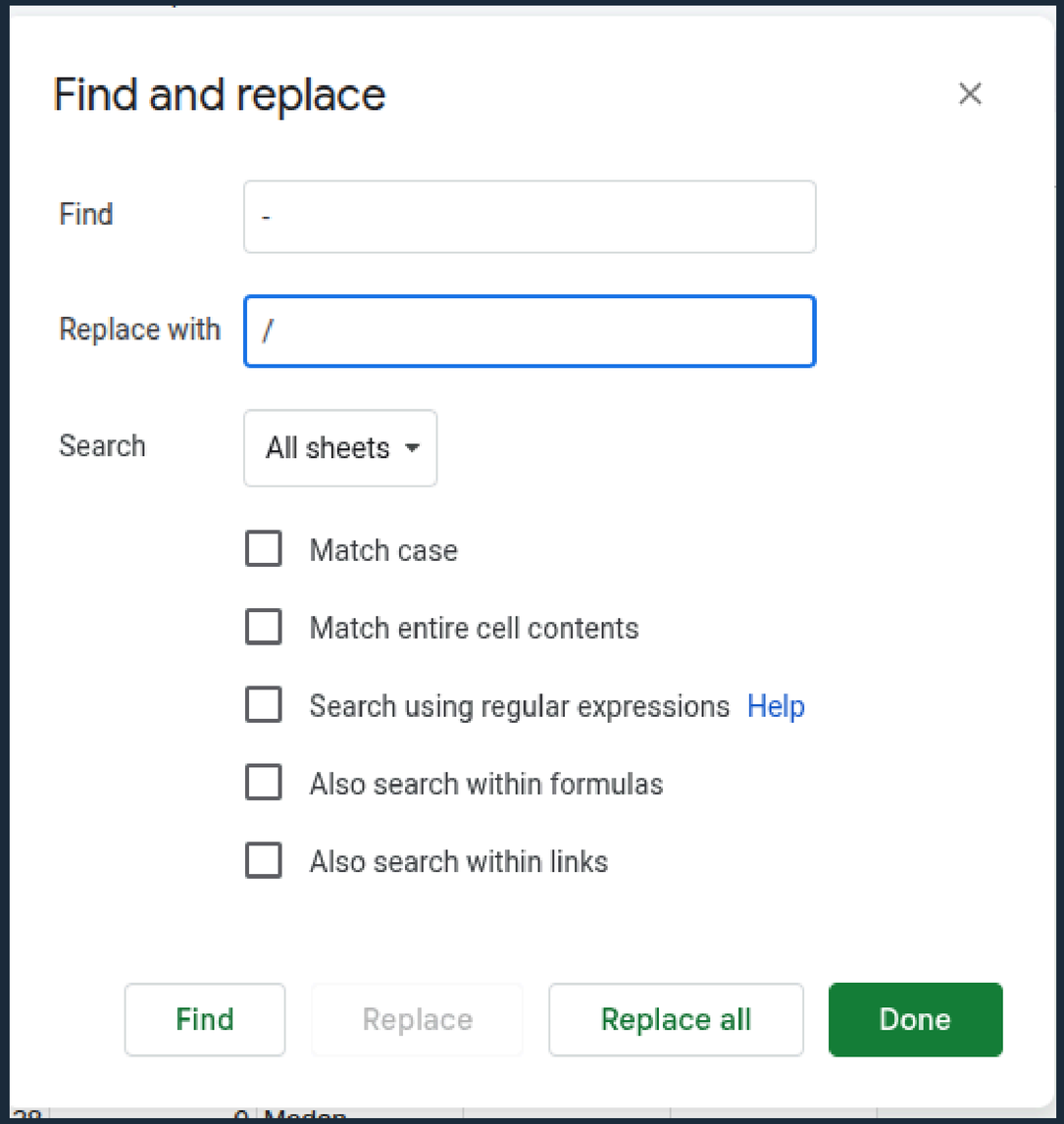
Jan → 1

Feb → 2

...

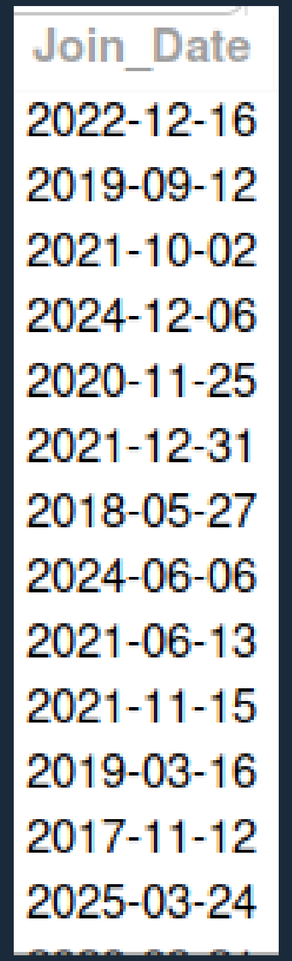
Dec → 12

In addition, replace ‘/’ with ‘-’ using the Find and Replace feature in Google Sheets.



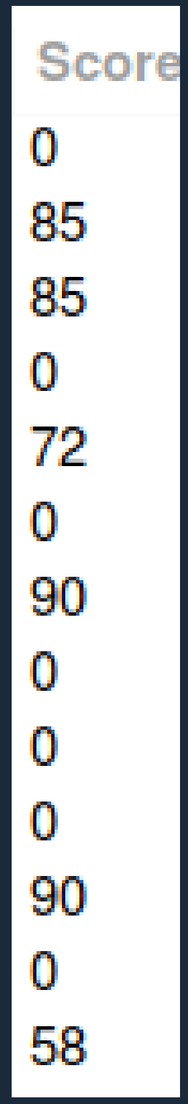
Date Format to YYYY-MM-DD

Third, I use MySQL Workbench as part of the data cleaning process. In this step, I run an UPDATE query that uses a CASE statement to standardize and format the date values. By applying this logic directly in SQL, the formatting process becomes significantly faster and more efficient compared to manual editing. This method ensures consistency across all rows and reduces the risk of errors during data transformation.



Make Empty Column table to zero

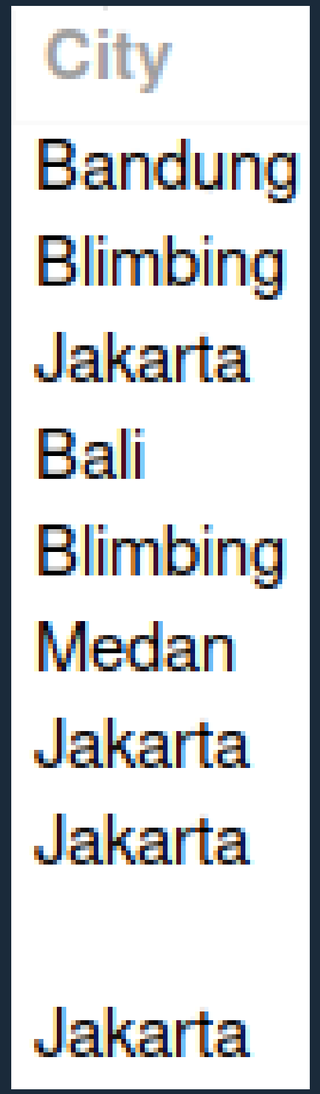
Converting empty fields to zero is a straightforward task. I apply an SQL UPDATE query to detect empty values and replace them with zero. Using SQL for this step ensures fast execution, consistency across the dataset, and minimizes the possibility of manual errors.



Change City because wrong typing

Berikut versi yang lebih panjang, jelas, dan grammarnya benar:
“I also correct text errors caused by typos. For example, ‘Jakrta’ is corrected to ‘Jakarta,’ and ‘Bandng’ is corrected to ‘Bandung.’
To fix these issues, I use an SQL UPDATE statement to identify the incorrect values and replace them with the correct spelling. This ensures that all city names are standardized and accurate across the entire dataset.”

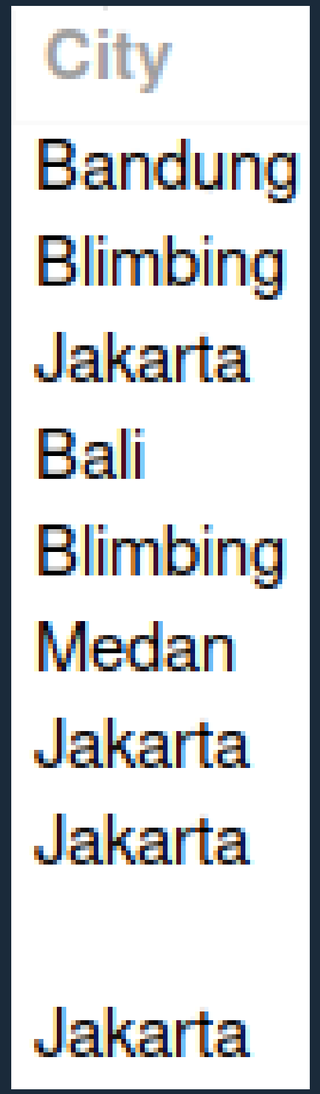
The image shows the result after the change.



Change City because wrong typing

Berikut versi yang lebih panjang, jelas, dan grammarnya benar:
“I also correct text errors caused by typos. For example, ‘Jakrta’ is corrected to ‘Jakarta,’ and ‘Bandng’ is corrected to ‘Bandung.’
To fix these issues, I use an SQL UPDATE statement to identify the incorrect values and replace them with the correct spelling. This ensures that all city names are standardized and accurate across the entire dataset.”

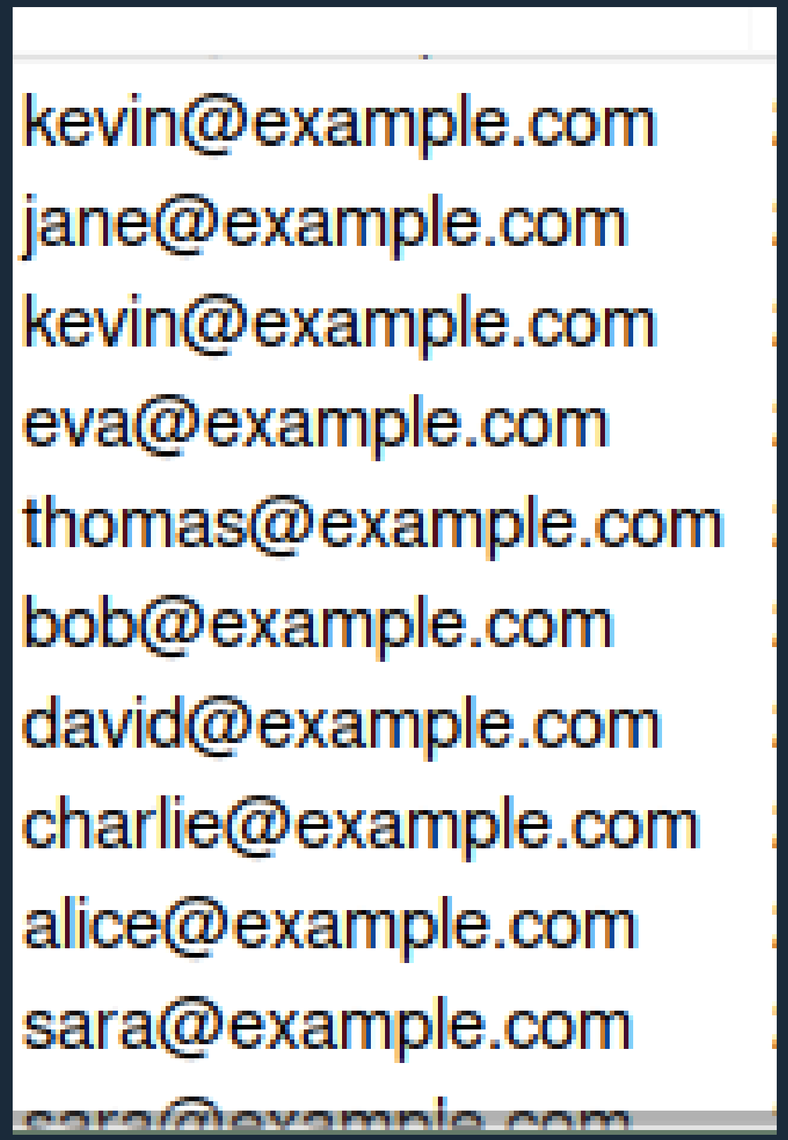
The image shows the result after the change.



Format Email

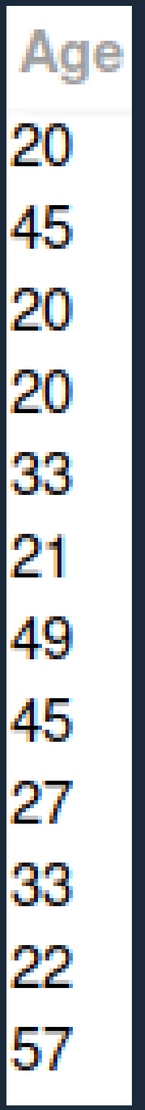
“In the email column, there are several different formats such as name@example and name[at]example.com. I need to correct these so that all emails use the standard format name@example.com.

To achieve this, I use an SQL UPDATE statement with REPLACE to convert the incorrect patterns into the correct email format. This ensures that all email addresses are consistent and properly standardized.”



Format Age

For the Age column, many values were empty, so I calculated the mode using SQL to fill in the missing entries. In addition, some age values contained unnecessary spaces, which made the format inconsistent. I corrected this by using an SQL UPDATE statement to clean the spacing and standardize the age values. This ensures that the Age column is complete, consistent, and properly formatted.



Age
20
45
20
20
33
21
49
45
27
33
22
57

Summary

SQL provides a fast and efficient way to process data. Beyond data cleaning, SQL is a powerful tool for filtering and retrieving information from tables. By leveraging well-constructed queries, I can standardize values, eliminate inconsistencies, and extract relevant subsets of data for subsequent analysis, significantly reducing manual effort and processing time.

Thank you

EMAIL

stevencandra797@gmail.com

LINK SOURCE DATA

https://github.com/stevencandra797/SQL-Dashboard/tree/main/Raw_Data